Nonparametric estimation of an extreme-value copula in arbitrary dimensions

Gordon Gudendorf[1], Johan Segers[1,*]

Institut de Statistique, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

Abstract

Inference on an extreme-value copula usually proceeds via its Pickands dependence function, which is a convex function on the unit simplex satisfying certain inequality constraints. In the setting of an iid random sample from a multivariate distribution with known margins and unknown extreme-value copula, an extension of the Capéraà–Fougères–Genest estimator was introduced by D. Zhang, M. T. Wells and L. Peng [Journal of Multivariate Analysis 99 (2008) 577–588]. The joint asymptotic distribution of the estimator as a random function on the simplex was not provided. Moreover, implementation of the estimator requires the choice of a number of weight functions on the simplex, the issue of their optimal selection being left unresolved.

A new, simplified representation of the CFG-estimator combined with standard empirical process theory provides the means to uncover its asymptotic distribution in the space of continuous, real-valued functions on the simplex. Moreover, the ordinary least-squares estimator of the intercept in a certain linear regression model provides an adaptive version of the CFG-estimator whose asymptotic behavior is the same as if the variance-minimizing weight functions were used. As illustrated in a simulation study, the gain in efficiency can be quite sizeable.

Key words: empirical process, linear regression, minimum-variance estimator, multivariate extreme-value distribution, ordinary least squares, Pickands dependence function, unit simplex
2010 MSC: 60F17, 62G32, 62H20

*Corresponding author
Email addresses: `gordon.gudendorf@uclouvain.be` (Gordon Gudendorf), `johan.segers@uclouvain.be` (Johan Segers)

[1]Research supported by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy) and by contract nr. 07/12/002 of the Projet d'Actions de Recherche Concertées of the Communauté française de Belgique, granted by the Académie universitaire Louvain.

1. Introduction

Let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})$, $i \in \{1, \ldots, n\}$, be iid random vectors from a p-variate, continuous distribution function F with multivariate extreme-value copula C: for $\boldsymbol{u} \in (0,1]^p \setminus \{(1, \ldots, 1)\}$, denoting the margins of F by $F_1, \ldots, F_p$,

$$C(\boldsymbol{u}) = P\big(F_1(X_{i1}) \leqslant u_1, \ldots, F_p(X_{ip}) \leqslant u_p\big) = \exp\{-|\boldsymbol{y}|\, A(\boldsymbol{y}/|\boldsymbol{y}|)\}$$
$$\text{where } y_j = -\log u_j \text{ and } |\boldsymbol{y}| = |y_1| + \cdots + |y_p|. \quad (1.1)$$

The function A, whose domain is $\Delta_p = \{\boldsymbol{w} \in [0,1]^p : w_1 + \cdots + w_p = 1\}$, is called the Pickands dependence function of C, after Pickands (1981).

Multivariate extreme-value copulas arise as the limits of copulas of vectors of component-wise maxima of independent random samples (Deheuvels, 1984; Galambos, 1987). As a consequence, they coincide with the class of copulas of multivariate extreme-value or max-stable distributions. Therefore, they provide models for dependence between extreme values that allow extrapolation beyond the support of the sample. It is then of interest to estimate the Pickands dependence function A.

A necessary condition for C in (1.1) to be a copula is that A is convex and satisfies $\max(w_1, \ldots, w_p) \leqslant A(\boldsymbol{w}) \leqslant 1$ for all $\boldsymbol{w} \in \Delta_p$; in the bivariate case, this is also sufficient. In general, A should admit an integral representation in terms of a spectral measure. Some other properties of Pickands dependence functions are studied in Obretenov (1991) and Falk and Reiss (2008). The upshot of all this is that the class of Pickands dependence functions is infinite-dimensional. This warrants the use of nonparametric methods.

Whereas most papers hitherto concentrated on the bivariate case, a nonparametric estimator for general multivariate Pickands dependence functions was introduced in Zhang et al. (2008). This estimator is in fact a multivariate generalization of the one by Capéraà–Fougères–Genest (Capéraà et al., 1997). The estimator was shown to be uniformly consistent and pointwise asymptotically normal. However, the joint asymptotic distribution of the estimator as a random function on Δ_p was not provided. Moreover, implementation of the estimator requires the choice of p weight functions λ_j on Δ_p, the issue of their optimal selection being left unresolved.

Using a simplified representation of the above-mentioned estimator, we are able to uncover its asymptotic distribution in the space $\mathscr{C}(\Delta_p)$ of continuous, real-valued functions on Δ_p. Moreover, we give explicit expressions for the weight functions λ_j that minimize the pointwise asymptotic variance of the estimator. These optimal weight functions depend on the unknown distribution. We show that the CFG-estimator with estimated variance-minimizing weight functions can be implemented as the intercept estimator in a certain linear regression model via ordinary least squares. The OLS-estimator is data-adaptive in the sense that the asymptotic distribution is the same as if the optimal weight functions were used. In a simulation study, the gain in efficiency is shown to be quite sizeable.

As in Zhang et al. (2008), the setting here is that of a random sample from a distribution whose margins are known and whose copula is an extreme-value

copula. It would be worthwhile to extend this to the case of unknown margins (Guillotte and Perron, 2008; Genest and Segers, 2009) and the case that the copula of F is merely in the domain of attraction of an extreme-value copula (Capéraà and Fougères, 2000; Einmahl and Segers, 2009).

The outline of our paper is as follows. The CFG-estimator is introduced in the next section, including its simplified representation and asymptotic distribution. The variance-minimizing weight functions are computed in Section 3 together with an adaptive estimator based on ordinary least squares in a linear regression framework. Section 4 reports on a simulation study. The proofs of the results in Sections 2 and 3 are deferred to Appendices A and B, respectively.

2. CFG-estimator and variants

Let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})$, $i \in \{1, \ldots, n\}$, be iid random vectors from a p-variate, continuous distribution function F with multivariate extreme-value copula C and Pickands dependence function A as in (1.1). Let $F_1, \ldots, F_p$ be the marginal distribution functions of F. Put $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})$ where

$$Y_{ij} = -\log F_j(X_{ij}) \tag{2.1}$$

for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, p\}$. The marginal distributions of the random variables Y_{ij} are standard exponential. The random vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_p$ are iid with common joint survivor function

$$P(Y_{i1} > y_1, \ldots, Y_{ip} > y_p) = C(e^{-y_1}, \ldots, e^{-y_p}) = \exp\{-|\boldsymbol{y}|\, A(\boldsymbol{y}/|\boldsymbol{y}|)\},$$

for $\boldsymbol{y} \in [0, \infty)^p \setminus \{(0, \ldots, 0)\}$, where $|\boldsymbol{y}| = |y_1| + \cdots + |y_p|$. Put

$$\xi_i(\boldsymbol{w}) = \bigwedge_{j=1}^{p} \frac{Y_{ij}}{w_j}, \qquad \boldsymbol{w} \in \Delta_p,\ i \in \{1, \ldots, n\}, \tag{2.2}$$

with '$\wedge$' denoting minimum and with the obvious convention for division by zero; in particular, $\xi_i(\boldsymbol{e}_j) = Y_{ij}$ for the p standard unit vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p$ in $\mathbb{R}^p$. For $\boldsymbol{w} \in \Delta_p$ and $x > 0$, we have

$$P\bigl(\xi_i(\boldsymbol{w}) > x\bigr) = P(Y_{i1} > w_1 x, \ldots, Y_{ip} > w_p x) = \exp\{-x\, A(\boldsymbol{w})\}. \tag{2.3}$$

Hence the random variables $\xi_1(\boldsymbol{w}), \ldots, \xi_n(\boldsymbol{w})$ constitute an independent random sample from the exponential distribution with mean $1/A(\boldsymbol{w})$. It follows that the distribution of $-\log \xi_i(\boldsymbol{w})$ is Gumbel with location parameter $\log A(\boldsymbol{w})$, whence

$$E[-\log \xi_i(\boldsymbol{w})] = \log A(\boldsymbol{w}) + \gamma, \tag{2.4}$$

the Euler–Mascheroni constant $\gamma = -\Gamma'(1) = 0.5772\ldots$ being the mean of the standard Gumbel distribution. This suggests the naive estimator

$$\log \hat{A}_n(\boldsymbol{w}) = -\frac{1}{n} \sum_{i=1}^{n} \log \xi_i(\boldsymbol{w}) - \gamma, \qquad \boldsymbol{w} \in \Delta_p. \tag{2.5}$$

The naive estimator is itself not a valid Pickands dependence function. For instance, it does not verify the vertex constraints $A(\boldsymbol{e}_j) = 1$ for all $j \in \{1, \ldots, p\}$. A simple way to at least remedy this defect is by putting

$$\log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) = \log \hat{A}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \log \hat{A}_n(\boldsymbol{e}_j), \qquad \boldsymbol{w} \in \Delta_p, \tag{2.6}$$

where $\lambda_1, \ldots, \lambda_p : \Delta_p \to \mathbb{R}$ are continuous functions verifying $\lambda_j(\boldsymbol{e}_k) = \delta_{jk}$ for all $j, k \in \{1, \ldots, p\}$. Continuity of the functions λ_j is assumed merely to ensure that the resulting estimator is a continuous function of $\boldsymbol{w}$ as well.

The superscript 'CFG' refers to the bivarate estimator by Capéraà–Fougères–Genest in Capéraà et al. (1997), generalized to the multivariate case in Zhang et al. (2008). Actually, the original definition in Zhang et al. (2008) is

$$\log \hat{A}_n^{\mathrm{ZWP}}(\boldsymbol{w}) = \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \int_0^{1-w_j} \frac{n^{-1} \sum_{i=1}^{n} \mathbf{1}\{Z_{ij}(\boldsymbol{w}) \leqslant z\} - z}{z(1-z)} \, dz, \tag{2.7}$$

where, with Y_{ij} as in (2.1),

$$Z_{ij}(\boldsymbol{w}) = \frac{\bigwedge_{k:k \neq j} \frac{Y_{ik}}{w_k}}{\frac{Y_{ij}}{1-w_j} + \bigwedge_{k:k \neq j} \frac{Y_{ik}}{w_k}}, \qquad \boldsymbol{w} \in \Delta_p.$$

Moreover, in (2.7), the weight functions λ_j are supposed to be nonnegative and to satisfy the additional constraint

$$\sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) = 1, \qquad \boldsymbol{w} \in \Delta_p. \tag{2.8}$$

However, if (2.8) holds, then actually the two estimators coincide, that is,

$$\hat{A}_n^{\mathrm{ZWP}}(\boldsymbol{w}) = \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}), \qquad \boldsymbol{w} \in \Delta_p. \tag{2.9}$$

The proof of (2.9) is essentially the same as the one in Segers (2007) for the bivariate case, the key being that the integrals in (2.7) can be solved:

$$\begin{aligned} &\int_0^{1-w_j} \frac{\mathbf{1}\{Z_{ij}(\boldsymbol{w}) \leqslant z\} - z}{z(1-z)} \, dz \\ &\quad = \log[1 - \{(1-w_j) \wedge Z_{ij}(\boldsymbol{w})\}] + \log(1-w_j) - \log\{(1-w_j) \wedge Z_{ij}(\boldsymbol{w})\} \\ &\qquad = \log Y_{ij} - \log \xi_i(\boldsymbol{w}). \end{aligned}$$

In our representation (2.6), however, there is no reason whatsoever to restrict the weight functions to satisfy (2.8).

The asymptotics of the naive estimator and the CFG-estimator follow from standard empirical process theory as presented for instance in van der Vaart and Wellner (1996) and van der Vaart (1998). Let $\mathscr{C}(\Delta_p)$ denote the Banach space of continuous functions from Δ_p into $\mathbb{R}$ equipped with the supremum norm. Convergence in distribution is denoted by the arrow '$\rightsquigarrow$'.

Proposition 2.1 (Naive estimator)**.** *Let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})$, $i \in \{1, \ldots, n\}$, be iid random variables from a p-variate, continuous distribution function F with multivariate extreme-value copula C and Pickands dependence function A. The naive estimator $\hat{A}_n$ in (2.5) satisfies*

$$\sup_{\boldsymbol{w} \in \Delta_p} |\hat{A}_n(\boldsymbol{w}) - A(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely}, \tag{2.10}$$

and in $\mathscr{C}(\Delta_p)$,

$$\sqrt{n}(\hat{A}_n - A) \rightsquigarrow A\,\zeta, \qquad n \to \infty, \tag{2.11}$$

where ζ is a centered Gaussian process with covariance function

$$\operatorname{cov}\bigl(\zeta(\boldsymbol{v}), \zeta(\boldsymbol{w})\bigr) = \operatorname{cov}\bigl(-\log \xi_i(\boldsymbol{v}),\, -\log \xi_i(\boldsymbol{w})\bigr), \qquad \boldsymbol{v}, \boldsymbol{w} \in \Delta_p, \tag{2.12}$$

with $\xi_i(\,\cdot\,)$ as in (2.2).

Theorem 2.2 (CFG-estimator)**.** *If, in addition to the assumptions in Proposition 2.1, the functions $\lambda_1, \ldots, \lambda_p : \Delta_p \to \mathbb{R}$ are continuous, then*

$$\sup_{\boldsymbol{w} \in \Delta_p} |\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - A(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely}, \tag{2.13}$$

and in $\mathscr{C}(\Delta_p)$,

$$\sqrt{n}(\hat{A}_n^{\mathrm{CFG}} - A) \rightsquigarrow A\,\eta, \qquad n \to \infty, \tag{2.14}$$

where η is a centered Gaussian process defined by

$$\eta(\boldsymbol{w}) = \zeta(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w})\,\zeta(\boldsymbol{e}_j), \qquad \boldsymbol{w} \in \Delta_p, \tag{2.15}$$

with ζ as in Proposition 2.1.

Remark 2.3 (Covariance function). The covariance function (2.12) can be expressed in terms of A as follows. An application of the identity $\log(x) = \int_0^\infty \{\mathbf{1}(s \leqslant x) - \mathbf{1}(s \leqslant 1)\}\, s^{-1}\, ds$ for $x \in (0, \infty)$ yields, by Fubini's theorem,

$$\begin{aligned}
&\operatorname{cov}\bigl(-\log \xi_i(\boldsymbol{v}),\, -\log \xi_i(\boldsymbol{w})\bigr) \\
&= \int_0^\infty \int_0^\infty \Bigl(P\bigl(\xi_i(\boldsymbol{v}) \geqslant s,\, \xi_i(\boldsymbol{w}) \geqslant t\bigr) - P\bigl(\xi_i(\boldsymbol{v}) \geqslant s\bigr)\, P\bigl(\xi_i(\boldsymbol{w}) \geqslant t\bigr) \Bigr) \frac{ds}{s}\, \frac{dt}{t} \\
&= \int_0^\infty \int_0^\infty [\exp\{-\ell((w_1 s) \vee (v_1 t), \ldots, (w_p s) \vee (v_p t))\} \\
&\qquad\qquad\qquad\qquad - \exp\{-s\, A(\boldsymbol{v})\} \exp\{-t\, A(\boldsymbol{w})\}] \frac{ds}{s}\, \frac{dt}{t}.
\end{aligned}$$

where $\ell(\boldsymbol{y}) = |\boldsymbol{y}|\, A(\boldsymbol{y}/|\boldsymbol{y}|)$ and $|\boldsymbol{y}| = |y_1| + \cdots + |y_p|$. Replacing A by any estimator of it results in an estimator of the covariance function. However, a more practical way to estimate this function is by the sample covariance of the pairs $(-\log \xi_i(\boldsymbol{v}), -\log \xi_i(\boldsymbol{w}))$; see also (the proof of) Theorem 3.2.

Remark 2.4 (Shape constraints). A further enhancement to the CFG-estimator is to replace it by the convex minorant of the function

$$\min[\max\{\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}), w_1, \ldots, w_p\}, 1], \qquad \boldsymbol{w} \in \Delta_p,$$

as in Deheuvels (1991) and Jiménez et al. (2001) for the bivariate case. Although the resulting estimator would be a convex function respecting the bounds $\max(w_1, \ldots, w_p) \leqslant A(\boldsymbol{w}) \leqslant 1$, in case $p \geqslant 3$ this would still not guarantee it to be a genuine Pickands dependence function. Still other ways to impose (some of) the shape restrictions are spline smoothing under constraints (Hall and Tajvidi, 2000), orthogonal projection (Fils-Villetard et al., 2008), or Bayesian nonparametrics (Guillotte and Perron, 2008).

Remark 2.5 (Pickands estimator). A different way to exploit the exponentiality of the random variables $\xi_i(\boldsymbol{w})$ in (2.3) would be via the Pickands estimator

$$\frac{1}{\hat{A}_n^{\mathrm{P}}(\boldsymbol{w})} = \frac{1}{n} \sum_{i=1}^{n} \xi_i(\boldsymbol{w})$$

as in Pickands (1981). To impose the vertex constraints $A(\boldsymbol{e}_j) = 1$, the techniques of Deheuvels (1991) or Hall and Tajvidi (2000) can be used, see Zhang et al. (2008, p. 578). In the bivariate case however, it is known that the resulting estimators are outperformed by the CFG-estimator $\hat{A}_n^{\mathrm{CFG}}$ (Segers, 2007; Genest and Segers, 2009). This is confirmed in the simulation study in Zhang et al. (2008, Section 3), as well as by our own simulations in Section 4. For this reason, we restrict attention here to the family of CFG-estimators.

3. The OLS-estimator

The question remains which weight functions λ_j to choose in the CFG-estimator (2.6). In Zhang et al. (2008), the choice $\lambda_j(\boldsymbol{w}) = w_j$ was recommended as a pragmatic one. The option of using variance-minimizing functions λ_j was mentioned but not carried out. By casting the estimation problem in a linear regression framework, we will obtain an estimator with the same asymptotic performance as the CFG-estimator with those optimal weights. In this section, we define the estimator and prove its consistency and asymptotic normality, both in the functional sense. In the next section, the gain in efficiency is assessed by means of simulations.

In view of Theorem 2.2, for each $\boldsymbol{w} \in \Delta_p$ we have

$$\sqrt{n}\big(\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - A(\boldsymbol{w})\big) \rightsquigarrow A(\boldsymbol{w})\, \eta(\boldsymbol{w}), \qquad n \to \infty,$$

where $\eta(\boldsymbol{w})$ is a zero-mean normal random variable. We will look for those $\lambda_j(\boldsymbol{w})$ that minimise the variance of $\eta(\boldsymbol{w})$. Let ζ be the Gaussian process on $\mathscr{C}(\Delta_p)$ in Proposition 2.1. For ease of notation, put

$$\boldsymbol{\lambda}(\boldsymbol{w}) = \big(\lambda_1(\boldsymbol{w}), \ldots, \lambda_p(\boldsymbol{w})\big)^\top, \qquad \boldsymbol{\zeta}(\boldsymbol{e}) = \big(\zeta(\boldsymbol{e}_1), \ldots, \zeta(\boldsymbol{e}_p)\big)^\top,$$

the symbol "$\top$" denoting matrix transposition. Then

$$\begin{aligned}\operatorname{var}\eta(\boldsymbol{w}) &= \operatorname{var}\big(\zeta(\boldsymbol{w}) - \boldsymbol{\lambda}(\boldsymbol{w})^\top \boldsymbol{\zeta}(\boldsymbol{e})\big) \\ &= \operatorname{var}\zeta(\boldsymbol{w}) - 2\,\boldsymbol{\lambda}(\boldsymbol{w})^\top E[\boldsymbol{\zeta}(\boldsymbol{e})\,\zeta(\boldsymbol{w})] + \boldsymbol{\lambda}(\boldsymbol{w})^\top E[\boldsymbol{\zeta}(\boldsymbol{e})\,\boldsymbol{\zeta}(\boldsymbol{e})^\top]\,\boldsymbol{\lambda}(\boldsymbol{w}).\end{aligned}$$

Note that

$$\Sigma = E[\boldsymbol{\zeta}(\boldsymbol{e})\,\boldsymbol{\zeta}(\boldsymbol{e})^\top] \tag{3.1}$$

is the covariance matrix of $(-\log\xi(\boldsymbol{e}_1), \ldots, -\log\xi(\boldsymbol{e}_p))^\top$. Provided this matrix is non-singular, $\operatorname{var}\eta(\boldsymbol{w})$ attains a unique global minimum for $\boldsymbol{\lambda}(\boldsymbol{w})$ equal to

$$\boldsymbol{\lambda}^{\mathrm{opt}}(\boldsymbol{w}) = \Sigma^{-1}\,E[\boldsymbol{\zeta}(\boldsymbol{e})\,\zeta(\boldsymbol{w})]. \tag{3.2}$$

With this choice of the weight functions, the variance of

$$\eta_{\mathrm{opt}}(\boldsymbol{w}) = \zeta(\boldsymbol{w}) - \boldsymbol{\lambda}^{\mathrm{opt}}(\boldsymbol{w})^\top\,\boldsymbol{\zeta}(\boldsymbol{e}) \tag{3.3}$$

is equal to

$$\operatorname{var}\eta_{\mathrm{opt}}(\boldsymbol{w}) = \operatorname{var}\zeta(\boldsymbol{w}) - E[\zeta(\boldsymbol{w})\,\boldsymbol{\zeta}(\boldsymbol{e})^\top]\,\Sigma^{-1}\,E[\boldsymbol{\zeta}(\boldsymbol{e})\,\zeta(\boldsymbol{w})]. \tag{3.4}$$

This variance is minimal over all possible choices of weight functions λ_j.

The optimal weight functions λ_j^{opt} in (3.2) depend on the unknown Pickands dependence function A. Fortunately, replacing these weight functions by uniformly consistent estimators $\hat\lambda_{n,j}$ is just as good asymptotically. For such estimated weight functions, define the adaptive CFG-estimator by

$$\log\hat A_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w}) = \log\hat A_n(\boldsymbol{w}) - \sum_{j=1}^p \hat\lambda_{n,j}(\boldsymbol{w})\,\log\hat A_n(\boldsymbol{e}_j), \tag{3.5}$$

Proposition 3.1 (Adaptive CFG-estimator)**.** *Assume that, in addition to the assumptions in Proposition 2.1, the matrix Σ in* (3.1) *is non-singular and $\hat\lambda_{n,j}$ are random elements in $\mathscr{C}(\Delta_p)$ such that, for every $j \in \{1, \ldots, p\}$,*

$$\sup_{\boldsymbol{w}\in\Delta_p} |\hat\lambda_{n,j}(\boldsymbol{w}) - \lambda_j^{\mathrm{opt}}(\boldsymbol{w})| \to 0, \qquad n\to\infty, \qquad \textit{almost surely},$$

with λ_j^{opt} as in (3.2)*. Then the adaptive CFG-estimator in* (3.5) *satisfies*

$$\sup_{\boldsymbol{w}\in\Delta_p} |\hat A_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w}) - A(\boldsymbol{w})| \to 0, \qquad n\to\infty, \qquad \textit{almost surely}, \tag{3.6}$$

and in $\mathscr{C}(\Delta_p)$,

$$\sqrt{n}(\hat A_{n,\mathrm{ad}}^{\mathrm{CFG}} - A) \rightsquigarrow A\,\eta_{\mathrm{opt}}, \qquad n\to\infty, \tag{3.7}$$

where η_{opt} is the zero-mean Gaussian process defined in (3.3)*.*

Finally we propose a particularly convenient way to implement the adaptive CFG-estimator in (3.5). For $\boldsymbol{w} \in \Delta_p$, let $\hat{\boldsymbol{\beta}}_n(\boldsymbol{w}) = (\hat{\beta}_{n,0}(\boldsymbol{w}), \ldots, \hat{\beta}_{n,p}(\boldsymbol{w}))^\top$ be the minimizer in $(b_0, \ldots, b_p)^\top$ of

$$\sum_{i=1}^{n} \Big(\big(-\log \xi_i(\boldsymbol{w}) - \gamma\big) - b_0 - \sum_{j=1}^{p} b_j \big(-\log \xi_i(\boldsymbol{e}_j) - \gamma\big) \Big)^2. \tag{3.8}$$

In words, $\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})$ is the ordinary least-squares (OLS) estimator of the vector of regression coefficients in a linear regression of the dependent variable $-\log \xi_i(\boldsymbol{w}) - \gamma$ upon the explanatory variables $-\log \xi_i(\boldsymbol{e}_j) - \gamma$, $j \in \{1, \ldots, p\}$. Define the OLS-estimator of A via the estimated intercept by

$$\log \hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w}) = \hat{\beta}_{n,0}(\boldsymbol{w}), \qquad \boldsymbol{w} \in \Delta_p.$$

Since the residuals

$$\hat{\epsilon}_{n,i}(\boldsymbol{w}) = \big(-\log \xi_i(\boldsymbol{w}) - \gamma\big) - \hat{\beta}_{n,0}(\boldsymbol{w}) - \sum_{j=1}^{p} \hat{\beta}_{n,j}(\boldsymbol{w}) \big(-\log \xi_i(\boldsymbol{e}_j) - \gamma\big)$$

verify $\sum_{i=1}^n \hat{\epsilon}_{n,i}(\boldsymbol{w}) = 0$, we have

$$\log \hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w}) = \hat{\beta}_{n,0}(\boldsymbol{w}) = \log \hat{A}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \hat{\beta}_{n,p}(\boldsymbol{w}) \, \log \hat{A}_n(\boldsymbol{e}_j), \tag{3.9}$$

that is, the OLS-estimator is equal to the adaptive CFG-estimator with estimated weights $\hat{\lambda}_{n,j}(\boldsymbol{w}) = \hat{\beta}_{n,j}(\boldsymbol{w})$. The variance of the (logarithm of the) OLS-estimator can be estimated by the sample variance of the residuals, properly corrected for the loss in number of degrees of freedom,

$$\hat{\sigma}^2_{n,\mathrm{OLS}}(\boldsymbol{w}) = \frac{1}{n-p-1} \sum_{i=1}^{n} \hat{\epsilon}^2_{n,i}(\boldsymbol{w}), \qquad \boldsymbol{w} \in \Delta_p. \tag{3.10}$$

Theorem 3.2 (OLS-estimator)**.** *Assume that, in addition to the assumptions in Proposition 2.1, the matrix* Σ *in* (3.1) *is non-singular. Then, with probability tending to one, the minimizer* $\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})$ *of* (3.8) *is uniquely defined and for* $j \in \{1, \ldots, p\}$,

$$\sup_{\boldsymbol{w} \in \Delta_p} |\hat{\beta}_{n,j}(\boldsymbol{w}) - \lambda_j^{\mathrm{opt}}(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely.} \tag{3.11}$$

As a consequence, the OLS-estimator in (3.9) *is uniformly consistent,*

$$\sup_{\boldsymbol{w} \in \Delta_p} |\hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w}) - A(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely,} \tag{3.12}$$

and in $\mathscr{C}(\Delta_p)$,

$$\sqrt{n}(\hat{A}_n^{\mathrm{OLS}} - A) \rightsquigarrow A\, \eta_{\mathrm{opt}}, \qquad n \to \infty, \tag{3.13}$$

where η_{opt} is the zero-mean Gaussian process defined in (3.3)*. In addition, the variance estimator in* (3.10) *satisfies*

$$\sup_{\boldsymbol{w}\in\Delta_p} |\hat{\sigma}^2_{n,\mathrm{OLS}}(\boldsymbol{w}) - \operatorname{var}\eta_{\mathrm{opt}}(\boldsymbol{w})| \to 0, \qquad n\to\infty, \qquad \textit{almost surely.}$$

Remark 3.3 (Non-singularity assumption). In the bivariate case, the assumption that the covariance matrix Σ in (3.1) is non-singular is equivalent to the assumption that the copula C is not the comonotone copula (Segers, 2007). We conjecture that in the general multivariate case, a necessary and sufficient condition for Σ to be non-singular is that none of the bivariate margins of C is equal to the comonotone copula.

4. Simulations

In order to investigate the finite-sample properties of the estimators discussed in the previous sections, we generated pseudo-random samples from trivariate extreme-value copulas of logistic type as presented in Tawn (1990):

$$\begin{aligned} A(\boldsymbol{w}) = (\theta^r w_1^r + \phi^r w_2^r)^{1/r} + (\theta^r w_2^r + \phi^r w_3^r)^{1/r} + (\theta^r w_3^r + \phi^r w_1^r)^{1/r} \\ + \psi(w_1^r + w_2^r + w_3^r)^{1/r} + 1 - \theta - \phi - \psi, \qquad \boldsymbol{w}\in\Delta_p, \end{aligned} \tag{4.1}$$

for $(r,\theta,\phi,\psi) \in [1,\infty)\times[0,1]^3$. To facilitate comparisons, we opted for the same parameter values as chosen in Zhang et al. (2008): a symmetric case, $(r,\theta,\phi,\psi) = (3,0,0,1)$, and an asymmetric one, $(r,\theta,\phi,\psi) = (6,0.6,0.3,0)$. For each case 10 000 samples were generated of size $n \in \{50,100,200\}$ using the simulation algorithms in Stephenson (2003) and implemented in the R-package evd (Stephenson, 2002).

Four estimators were compared: the CFG-estimator $\hat{A}_n^{\mathrm{CFG}}$ with weight functions $\lambda_j(\boldsymbol{w}) = w_j$ (as recommended in Zhang et al., 2008), the OLS-estimator $\hat{A}_n^{\mathrm{OLS}}$ in (3.9), and the enhanced versions of the original Pickands estimator due to Deheuvels (1991) and Hall and Tajvidi (2000) as presented in Zhang et al. (2008). To visualize the performances of the estimators, we plotted their biases and mean squared errors along the line $\{\boldsymbol{w}\in\Delta_p : w_1 = w_2\}$; see Figures 1 and 2 for the symmetric and asymmetric logistic dependence functions respectively.

In accordance to the theory, the OLS-estimator is in virtually all cases considered more efficient than the CFG-estimator. Moreover, our simulations confirm the findings in Zhang et al. (2008) that the CFG-estimator is typically more efficient than the ones of Deheuvels and Hall–Tajvidi. Note that the finite-sample bias of the OLS-estimator is somewhat larger than for the other estimators. However, thanks to its minimum-variance property it ends up as an overall winner in terms of mean squared error.

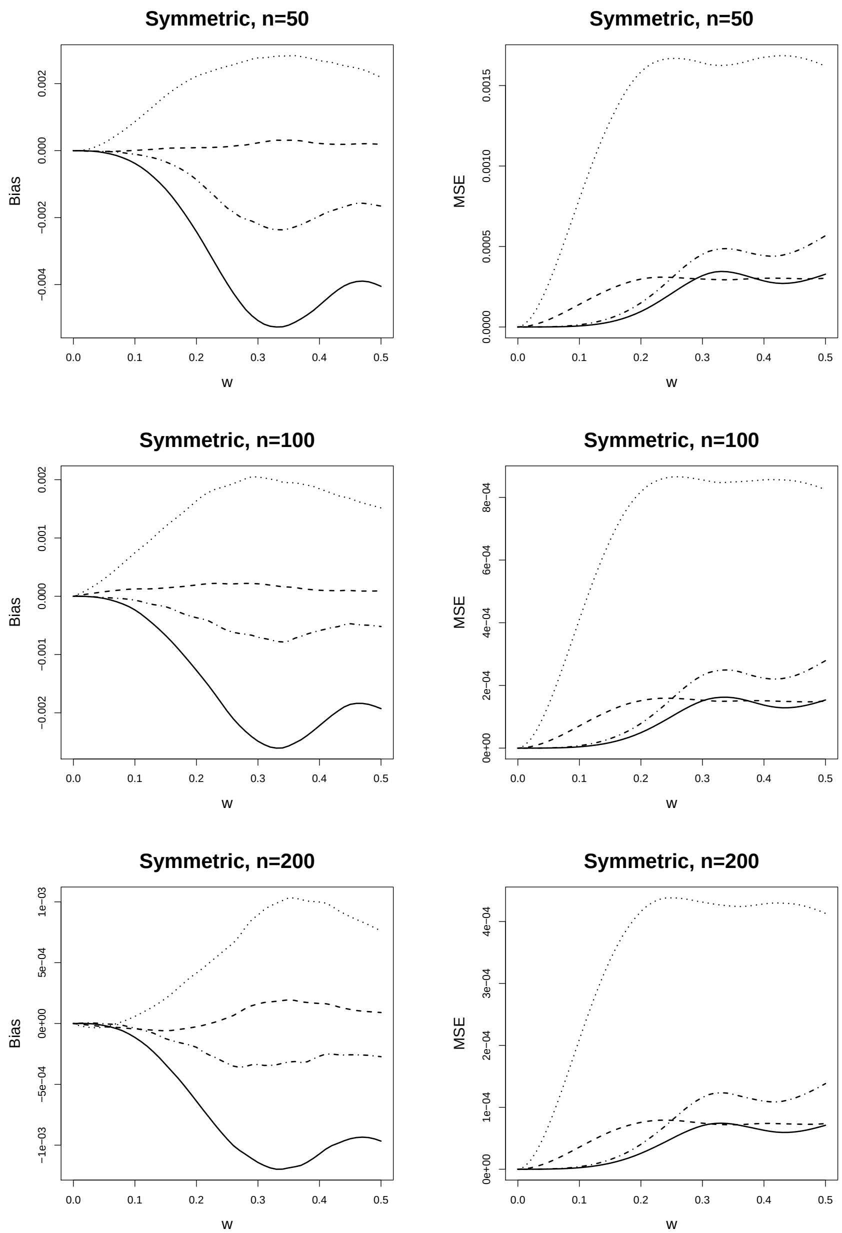

Figure 1: Biases (left) and mean squared errors (right) of $\hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w})$ (solid), $\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w})$ (dashed), $\hat{A}_n^{\mathrm{HT}}(\boldsymbol{w})$ (dash-dotted) and $\hat{A}_n^{\mathrm{D}}(\boldsymbol{w})$ (dotted) along the line $w_1 = w_2$ for 10 000 samples of size $n \in \{50, 100, 200\}$ from the trivariate extreme-value copula C with symmetric logistic dependence function $A(\boldsymbol{w}) = (w_1^r + w_2^r + w_3^r)^{1/r}$ at $r = 3$.

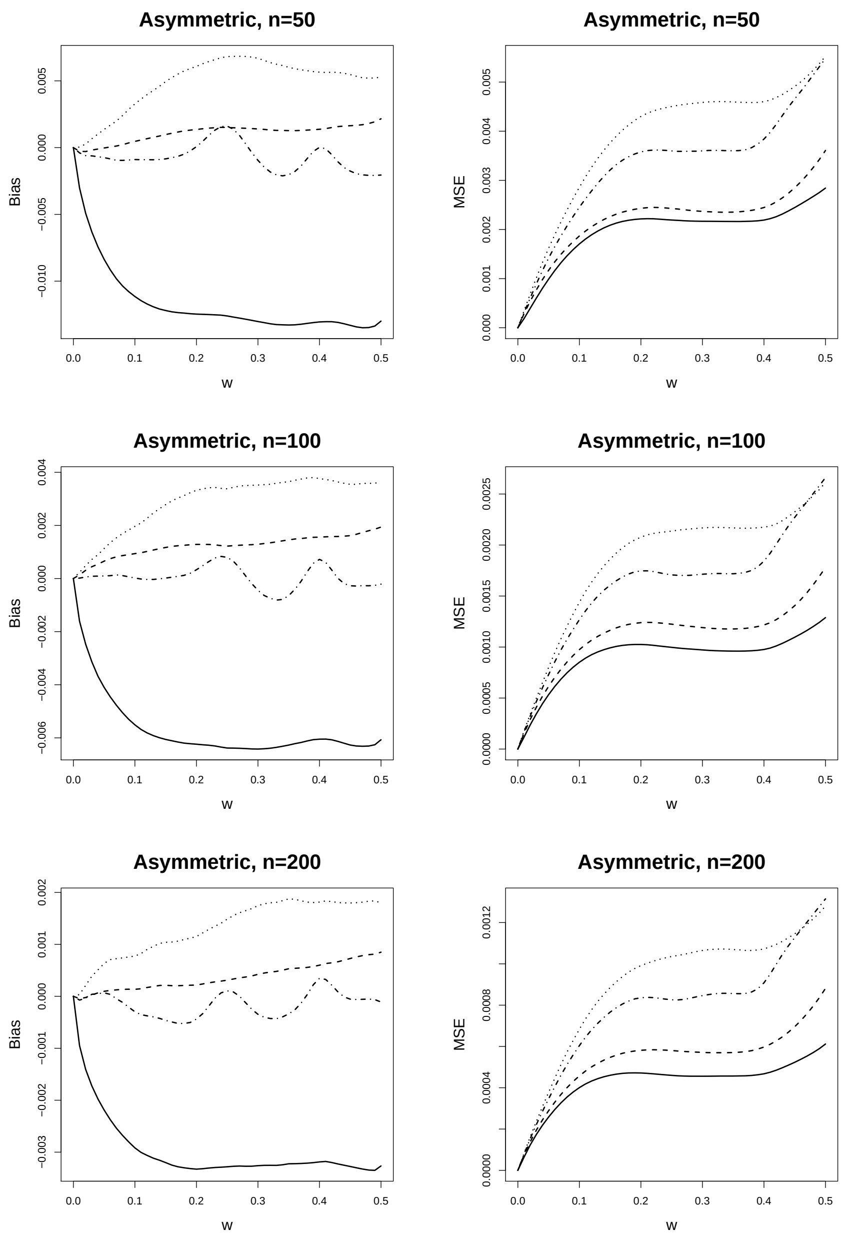

Figure 2: Biases (left) and mean squared errors (right) of $\hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w})$ (solid), $\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w})$ (dashed), $\hat{A}_n^{\mathrm{HT}}(\boldsymbol{w})$ (dash-dotted) and $\hat{A}_n^{\mathrm{D}}(\boldsymbol{w})$ (dotted) along the line $w_1 = w_2$ for 10 000 samples of size $n \in \{50, 100, 200\}$ from the trivariate extreme-value copula C with asymmetric logistic dependence function A in (4.1) for $(r, \theta, \phi, \psi) = (6, 0.6, 0.3, 0)$.

Appendix A. Proofs for Section 2

Proof of Proposition 2.1. For $\boldsymbol{w} \in \Delta_p$, define $f_{\boldsymbol{w}} : (0,\infty)^p \to \mathbb{R}$ by

$$f_{\boldsymbol{w}}(\boldsymbol{y}) = -\log\left(\bigwedge_{j=1}^{p} \frac{y_j}{w_j}\right) - \gamma, \qquad \boldsymbol{y} \in (0,\infty)^p. \tag{A.1}$$

We can write

$$\log \hat{A}_n(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n} f_{\boldsymbol{w}}(\boldsymbol{Y}_i).$$

Consider the function class $\mathscr{F} = \{f_{\boldsymbol{w}} : \boldsymbol{w} \in \Delta_p\}$. We will show that $\mathscr{F}$ is P-Donsker and therefore also P-Glivenko–Cantelli, where P denotes the common probability distribution on $(0,\infty)^p$ of the random vectors $\boldsymbol{Y}_i$. According to Theorem 2.6.8 in van der Vaart and Wellner (1996) and the proof thereof, we need to verify that $\mathscr{F}$ is a pointwise separable Vapnik–Červonenkis-class (VC-class) that admits an envelope function with a finite second moment under P. Pointwise separability follows from the fact that the map $\boldsymbol{w} \mapsto f_{\boldsymbol{w}}(\boldsymbol{y})$ is continuous in $\boldsymbol{w} \in \Delta_p$ for each $\boldsymbol{y} \in (0,\infty)^p$. The VC-property can be established by repeated applications of Lemmas 2.6.15 and 2.6.18, items (i) and (viii), in van der Vaart and Wellner (1996). Finally, the readily established bound

$$\left|\log \bigwedge_{j=1}^{p} \frac{y_j}{w_j}\right| \leqslant \max\left\{\left|\log \bigwedge_{j=1}^{p} y_j\right|, \log(p) + \sum_{j=1}^{p} |\log y_j|\right\} \tag{A.2}$$

yields an envelope function of $\mathscr{F}$ all of whose moments are finite under P. Observe that the distribution of $\bigwedge_{j=1}^{p} Y_{ij}$ is Exponential with mean equal to $\{p\,A(1/p,\ldots,1/p)\}^{-1} \in [1/p, 1]$.

From the fact that $\mathscr{F}$ is P-Glivenko–Cantelli it follows that

$$\begin{aligned}&\sup_{\boldsymbol{w}\in\Delta_p} |\log \hat{A}_n(\boldsymbol{w}) - \log A(\boldsymbol{w})| \\ &\quad = \sup_{\boldsymbol{w}\in\Delta_p}\left|\frac{1}{n}\sum_{i=1}^{n} f_{\boldsymbol{w}}(\boldsymbol{Y}_i) - E[f_{\boldsymbol{w}}(\boldsymbol{Y})]\right| \to 0, \qquad n\to\infty, \qquad \text{almost surely.}\end{aligned}$$

(Here, we dropped a subscript i for convenience.) Continuity of the map $\exp : \mathscr{C}(\Delta_p) \to \mathscr{C}(\Delta_p) : f \mapsto \exp(f)$ yields uniform consistency as in (2.10).

Moreover, the P-Donsker property entails

$$\sqrt{n}(\log \hat{A}_n^{\mathrm{CFG}} - \log A) \rightsquigarrow \zeta, \qquad n \to \infty, \tag{A.3}$$

in the space $\ell^\infty(\Delta_p)$ of bounded functions from Δ_p into $\mathbb{R}$ equipped with the topology of uniform convergence, where we identified $\mathscr{F}$ with Δ_p. The process ζ is zero-mean Gaussian with covariance function given in (2.12). The sample paths of the limit process ζ are continuous with respect to the standard deviation (semi-)metric ρ on Δ_p defined by

$$\rho(\boldsymbol{v}, \boldsymbol{w}) = [\mathrm{var}\{f_{\boldsymbol{v}}(\boldsymbol{Y}) - f_{\boldsymbol{w}}(\boldsymbol{Y})\}]^{1/2}, \qquad \boldsymbol{v}, \boldsymbol{w} \in \Delta_p.$$

If $\lim_{n\to\infty} \boldsymbol{v}_n = \boldsymbol{v}$ in Δ_p according to the Euclidean metric, then by continuity of $f_{\boldsymbol{w}}(\boldsymbol{y})$ in $\boldsymbol{w}$ and by uniform integrability, also $\lim_{n\to\infty} \rho(\boldsymbol{v}_n, \boldsymbol{v}) = 0$. (Uniform integrability is checked by using the bound in (A.2).) It follows that the trajectories of ζ are also continuous with respect to the Euclidean metric on Δ_p, that is, ζ actually takes its values in $\mathscr{C}(\Delta_p)$. As the trajectories of the left-hand side in (A.3) are continuous too, the convergence in (A.3) takes place not only $\ell^\infty(\Delta_p)$ but also in $\mathscr{C}(\Delta_p)$.

The convergence in (2.11) follows from the Hadamard-differentiability of the map $\exp : \mathscr{C}(\Delta_p) \to \mathscr{C}(\Delta_p) : f \mapsto \exp f$ and the functional delta-method (van der Vaart and Wellner, 1996, Section 3.9). $\square$

Proof of Theorem 2.2. Uniform consistency of $\hat{A}_n^{\mathrm{CFG}}$ in (2.13) follows from uniform consistency of $\hat{A}_n$ in (2.10) and the fact that the functions λ_j are continuous, hence bounded.

To show (2.14), define $L : \mathscr{C}(\Delta_p) \to \mathscr{C}(\Delta_p)$ by

$$Lf(\boldsymbol{w}) = f(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w})\, f(\boldsymbol{e}_j)$$

for $f \in \mathscr{C}(\Delta_p)$ and $\boldsymbol{w} \in \Delta_p$. The operator L is linear and bounded. We have $\log \hat{A}_n^{\mathrm{CFG}} = L(\log \hat{A}_n)$. Moreover, as $A(\boldsymbol{e}_j) = 1$ for all $j \in \{1, \ldots, p\}$, also $L(\log A) = \log A$. We find

$$\sqrt{n}(\log \hat{A}_n^{\mathrm{CFG}} - \log A) = L\big(\sqrt{n}(\log \hat{A}_n - \log A)\big) \rightsquigarrow L\zeta = \eta, \qquad n \to \infty.$$

The weak convergence in (2.14) follows from the functional delta-method (van der Vaart and Wellner, 1996, Section 3.9). The representation $\eta = L\zeta$ coincides with (2.15). $\square$

Appendix B. Proofs for Section 3

Proof of Proposition 3.1. If the optimal weight functions λ_j^{opt} were known, we could consider the optimal CFG-estimator

$$\log \hat{A}_{n,\mathrm{opt}}^{\mathrm{CFG}}(\boldsymbol{w}) = \log \hat{A}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j^{\mathrm{opt}}(\boldsymbol{w})\, \log \hat{A}_n(\boldsymbol{e}_j), \qquad \boldsymbol{w} \in \Delta_p.$$

By Theorem 2.2, the optimal CFG-estimator is uniformly consistent (2.13) and is asymptotically normal in the sense of (2.14) with $\eta = \eta_{\mathrm{opt}}$. Now

$$|\log \hat{A}_{n,\mathrm{opt}}^{\mathrm{CFG}}(\boldsymbol{w}) - \log \hat{A}_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w})| \leqslant \sum_{j=1}^{p} |\hat{\lambda}_{n,j}(\boldsymbol{w}) - \lambda_j^{\mathrm{opt}}(\boldsymbol{w})|\, |\log \hat{A}_n(\boldsymbol{e}_j)|.$$

By uniform consistency of $\hat{\lambda}_{n,j}$ and asymptotic normality of $\sqrt{n}\,\log \hat{A}_n(\boldsymbol{e}_j)$, we obtain, as $n \to \infty$,

$$\sup_{\boldsymbol{w}\in\Delta_p} |\log \hat{A}_{n,\mathrm{opt}}^{\mathrm{CFG}}(\boldsymbol{w}) - \log \hat{A}_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w})| \to 0, \qquad \text{almost surely,}$$

$$\sup_{\boldsymbol{w}\in\Delta_p} \sqrt{n}\, |\log \hat{A}_{n,\mathrm{opt}}^{\mathrm{CFG}}(\boldsymbol{w}) - \log \hat{A}_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w})| \rightsquigarrow 0.$$

As a consequence, the adaptive CFG-estimator is uniformly consistent (3.6) and asymptotically normal (3.7). □

Proof of Theorem 3.2. In analogy to the linear regression framework, define the $n \times (p+1)$ matrix

$$\boldsymbol{X} = \begin{pmatrix} 1 & -\log \xi_1(\boldsymbol{e}_1) - \gamma & \dots & \log \xi_1(\boldsymbol{e}_p) - \gamma \\ \dots & \dots & \dots & \dots \\ 1 & -\log \xi_n(\boldsymbol{e}_1) - \gamma & \dots & \log \xi_n(\boldsymbol{e}_p) - \gamma \end{pmatrix}$$

and the $n \times 1$ vector

$$\boldsymbol{Y}(\boldsymbol{w}) = \bigl(-\log \xi_1(\boldsymbol{w}) - \gamma, \dots, -\log \xi_n(\boldsymbol{w}) - \gamma\bigr)^\top, \qquad \boldsymbol{w} \in \Delta_p.$$

(No confusion should arise between this $\boldsymbol{Y}(\boldsymbol{w})$ and the random vectors $\boldsymbol{Y}_i$ in (2.1).) Provided the matrix $\boldsymbol{X}^\top \boldsymbol{X}$ is non-singular, the OLS-estimator $\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})$ is given by

$$\hat{\boldsymbol{\beta}}_n(\boldsymbol{w}) = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}(\boldsymbol{w}).$$

Recall the functions $f_{\boldsymbol{w}}$ in (A.1). For $\boldsymbol{v}, \boldsymbol{w} \in \Delta_p$, define $g_{\boldsymbol{v},\boldsymbol{w}} : (0,\infty)^p \to \mathbb{R}$ by

$$g_{\boldsymbol{v},\boldsymbol{w}}(\boldsymbol{y}) = f_{\boldsymbol{v}}(\boldsymbol{y})\, f_{\boldsymbol{w}}(\boldsymbol{y}), \qquad \boldsymbol{y} \in (0,\infty)^p.$$

By (A.2) and by Example 2.10.23 in van der Vaart and Wellner (1996), the function class $\{g_{\boldsymbol{v},\boldsymbol{w}} : \boldsymbol{v}, \boldsymbol{w} \in \Delta_p\}$ is P-Donsker and thus P-Glivenko–Cantelli, where P is the common distribution on $(0,\infty)^p$ of the random vectors $\boldsymbol{Y}_i$. It follows that, almost surely as $n \to \infty$,

$$\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X} \to \begin{pmatrix} 1 & 0 \\ 0 & \Sigma \end{pmatrix}, \tag{B.1}$$

$$\sup_{\boldsymbol{w} \in \Delta_p} \left| \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{Y}(\boldsymbol{w}) - \begin{pmatrix} \log A(\boldsymbol{w}) \\ E[\boldsymbol{\zeta}(\boldsymbol{e}) \zeta(\boldsymbol{w})] \end{pmatrix} \right| \to 0, \tag{B.2}$$

As Σ is non-singular, we have

$$\begin{pmatrix} 1 & 0 \\ 0 & \Sigma \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \Sigma^{-1} \end{pmatrix},$$

while $\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X}$ is with probability tending to one a non-singular matrix too. We find, almost surely and uniformly in $\boldsymbol{w} \in \Delta_p$,

$$\begin{aligned} \hat{\boldsymbol{\beta}}_n(\boldsymbol{w}) &= \left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X} \right)^{-1} \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{Y}(\boldsymbol{w}) \\ &\to \begin{pmatrix} 1 & 0 \\ 0 & \Sigma^{-1} \end{pmatrix} \begin{pmatrix} \log A(\boldsymbol{w}) \\ E[\boldsymbol{\zeta}(\boldsymbol{e}) \zeta(\boldsymbol{w})] \end{pmatrix} = \begin{pmatrix} \log A(\boldsymbol{w}) \\ \boldsymbol{\lambda}^{\mathrm{opt}}(\boldsymbol{w}) \end{pmatrix}, \qquad n \to \infty. \end{aligned}$$

Equation (3.11) follows. Proposition 3.1 and equation (3.9) then yield equations (3.12) and (3.13).

Finally, for the estimation of the variance, note that it does not matter asymptotically if we divide by n or by $n-p-1$. Elementary calculations yield

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_{n,i}^2(\boldsymbol{w}) &= \frac{1}{n}\big(\boldsymbol{Y}(\boldsymbol{w})-\boldsymbol{X}\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})\big)^\top\big(\boldsymbol{Y}(\boldsymbol{w})-\boldsymbol{X}\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})\big)\\ &= \frac{1}{n}\boldsymbol{Y}(\boldsymbol{w})^\top\boldsymbol{Y}(\boldsymbol{w}) - \left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{Y}(\boldsymbol{w})\right)^\top\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{Y}(\boldsymbol{w}).\end{aligned}$$

The Glivenko–Cantelli property yields, almost surely and uniformly in $\boldsymbol{w}\in\Delta_p$,

$$\begin{aligned}\frac{1}{n}\boldsymbol{Y}(\boldsymbol{w})^\top\boldsymbol{Y}(\boldsymbol{w}) &= \frac{1}{n}\sum_{i=1}^{n}\big(-\log\xi_i(\boldsymbol{w})-\gamma\big)^2\\ &\to E\Big[\big(-\log\xi_i(\boldsymbol{w})-\gamma\big)^2\Big] = \operatorname{var}\zeta(\boldsymbol{w}) + \big(\log A(\boldsymbol{w})\big)^2, \qquad n\to\infty.\end{aligned}$$

In combination with (B.1) and (B.2), we obtain that $n^{-1}\sum_{i=1}^n\hat{\epsilon}_{n,i}^2(\boldsymbol{w})$ converges almost surely and uniformly in $\boldsymbol{w}\in\Delta_p$ to

$$\begin{aligned}&\operatorname{var}\zeta(\boldsymbol{w}) + \big(\log A(\boldsymbol{w})\big)^2 - \begin{pmatrix}\log A(\boldsymbol{w})\\ E[\boldsymbol{\zeta}(\boldsymbol{e})\zeta(\boldsymbol{w})]\end{pmatrix}^\top\begin{pmatrix}1 & 0\\ 0 & \Sigma^{-1}\end{pmatrix}\begin{pmatrix}\log A(\boldsymbol{w})\\ E[\boldsymbol{\zeta}(\boldsymbol{e})\zeta(\boldsymbol{w})]\end{pmatrix}\\ &\qquad = \operatorname{var}\zeta(\boldsymbol{w}) - E[\boldsymbol{\zeta}(\boldsymbol{e})^\top\zeta(\boldsymbol{w})]\,\Sigma^{-1}\,E[\boldsymbol{\zeta}(\boldsymbol{e})\zeta(\boldsymbol{w})],\end{aligned}$$

which by (3.4) is equal to $\operatorname{var}\eta_{\mathrm{opt}}(\boldsymbol{w})$. $\square$

References

J. Pickands, Multivariate extreme value distributions, in: Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981), vol. 49, 859–878, 894–902, with a discussion, 1981.

P. Deheuvels, Probabilistic aspects of multivariate extremes, in: J. Tiago de Oliveira (Ed.), Statistical extremes and applications, Reidel, 117–130, 1984.

J. Galambos, The asymptotic theory of extreme order statistics, Krieger, Melbourne, Florida, 2nd edn., 1987.

A. Obretenov, On the dependence function of Sibuya in multivariate extreme value theory, Journal of Multivariate Analysis 36 (1) (1991) 35–43.

M. Falk, R. Reiss, On Pickands coordinates in arbitrary dimensions, Journal of Multivariate Analysis 92 (2) (2008) 426–453.

D. Zhang, M. T. Wells, L. Peng, Nonparametric estimation of the dependence function for a multivariate extreme value distribution, Journal of Multivariate Analysis 99 (4) (2008) 577–588.

P. Capéraà, A.-L. Fougères, C. Genest, A nonparametric estimation procedure for bivariate extreme value copulas, Biometrika 84 (1997) 567–577.

S. Guillotte, F. Perron, A Bayesian estimator for the dependence function of a bivariate extreme-value distribution, The Canadian Journal of Statistics 36 (3) (2008) 383–396.

C. Genest, J. Segers, Rank-based inference for bivariate extreme-value copulas, Annals of Statistics 37 (5B) (2009) 2990–3022.

P. Capéraà, A.-L. Fougères, Estimation of a bivariate extreme value distribution, Extremes 3 (2000) 311–329.

J. H. J. Einmahl, J. Segers, Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution, The Annals of Statistics 37 (5B) (2009) 2953–2989.

J. Segers, Non-parametric inference for bivariate extreme-value copulas, in: M. Ahsanulah, S. Kirmani (Eds.), Extreme Value Distributions, chap. 9, Nova Science Publishers, Inc., 181–203, older version available as CentER DP 2004-91, Tilburg University, 2007.

A. W. van der Vaart, J. A. Wellner, Weak Convergence and Empirical Processes, Cambridge Series in Statistical and Probabilistic Mathematics, Springer, New York, 1996.

A. W. van der Vaart, Asymptotic statistics, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, 1998.

P. Deheuvels, On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions, Statistics & Probability Letters 12 (5) (1991) 429–439.

J. R. Jiménez, E. Villa-Diharce, M. Flores, Nonparametric estimation of the dependence function in bivariate extreme value distributions, Journal of Multivariate Analysis 76 (2) (2001) 159–191.

P. Hall, N. Tajvidi, Distribution and dependence-function estimation for bivariate extreme-value distributions, Bernoulli 6 (5) (2000) 835–844.

A. Fils-Villetard, A. Guillou, J. Segers, Projection estimators of Pickands dependence functions, The Canadian Journal of Statistics 36 (3) (2008) 369–382.

J. Tawn, Modelling Multivariate Extreme Vlaue Distributions, Biometrika 77 (2) (1990) 245–253.

A. G. Stephenson, Simulating multivariate extrme value analysis of logistic type, Extremes 6 (2003) 49–59.

A. G. Stephenson, evd: Extreme Value Distributions, R News 2 (2) (2002) June, URL `http://CRAN.R-project.org/doc/Rnews/`.